THE LINEARIZATION OF THE CENTRAL LIMIT OPERATOR IN FREE PROBABILITY THEORY

MICHAEL ANSHELEVICH

ABSTRACT. We interpret the Central Limit Theorem as a fixed point theorem for a certain operator, and consider the problem of linearizing this operator. In classical as well as in free probability theory [VDN92], we consider two methods giving such a linearization, and interpret the result as a weak form of the CLT. In the classical case the analysis involves dilation operators; in the free case more general composition operators appear.

1. INTRODUCTION

The ubiquity of the normal distribution as indicated by the Central Limit Theorem (CLT) is a somewhat mysterious result. One of the possible explanations for it is an interpretation of the CLT as a fixed-point theorem; for a plethora of approaches see [Tro59, Gol76, HW84, Bar86, Swa91, Sin92]. The starting point for this analysis is the following weak form of the CLT for independent identically distributed random variables. Let the operator T be defined on probability measures on $\mathbb{R}$ by $T\mu = (\mu * \mu) \circ S_{\frac{1}{\sqrt{2}}}$. Here $*$ is the convolution, and S_r (for "scaling") is the dilation operator, $d(\mu \circ S_r)(x) = d\mu(r^{-1}x)$. We call this operator T the *central limit* operator. The theorem follows from the CLT, and is well known.

Theorem. *The fixed points of the operator T are the scaled normal distributions $\chi \circ S_t$. If μ is a probability measure with zero mean and unit variance, then the iterations $T^n\mu$ weakly converge to χ.*

In this approach, the starting point for the study of the CLT is the investigation of the operator T. This operator is clearly non-linear, and as the first approximation we consider the properties of the linearization of this operator [Sin92]. These properties are of course well-known (although we have not found adequate references; but see [Sin92] and also [Sin76]). However, there is now a different version of probability theory, with its own CLT, which has not been investigated to the same degree. This is the free probability theory of Voiculescu (for an introduction, see e.g. [VDN92]), which, in particular, turns out to describe the behavior of certain large random matrices. In this theory the notion of (commutative) independence is replaced by the notion of free independence (for the operator-theoretic definition and the motivation behind it see e.g. [VDN92]). Now the classical convolution

can be defined in terms of independence as follows: $\mu * \nu$ is the distribution of the sum of two independent random variables with distributions μ and ν (and it is a theorem that the distribution of the sum of independent random variables depends only on the distributions of the summands). Correspondingly, in free probability theory one defines the (additive) free convolution of measures $\mu \boxplus \nu$ as the distribution of the sum of two freely independent random variables with distributions μ and ν [Voi85, Voi86, BV93, VDN92] (and the above comment applies). Thus the *free central limit operator* is $T(\mu) = (\mu \boxplus \mu) \circ S_{\frac{1}{\sqrt{2}}}$. One of the main technical differences between classical and free theories of probability is that the operator of convolution with a given measure is linear, while the operator of taking a free convolution with a given measure is highly non-linear. However, the above operator T is non-linear even in the classical case, and thus one can expect similarities between linearizations of the classical and free versions of this operator.

We have two somewhat different approaches at our disposal. The original one, initiated and largely developed by Voiculescu [Voi86, BV93, VDN92] (see also [Maa92]), is to define a certain operation, called the *R-transform*, on the space of analytic functions, which linearizes the additive free convolution. Thus this operation is an analogue of the logarithm of the Fourier transform in the classical case. Another approach, due to Speicher [Spe90] and developed, among others, by Speicher and Nica [Spe94, Nic95], is to use a certain analogue of the classical combinatorial moment-cumulant formula. This approach is somewhat less general, but the parallel with the classical situation is more explicit.

In what follows we want to indicate the parallels between the classical and the free case. Therefore, whenever appropriate, we will use the same notation for both cases. The situations where (important) differences between the two theories arise will also be indicated.

Acknowledgments. We would like to thank Prof. D.-V. Voiculescu for suggesting the problem as well as many helpful discussions. We would also like to thank Prof. N.G. Makarov for some suggestions.

2. THE COMBINATORIAL APPROACH

2.1. **Notation.** Let μ be a probability measure. We denote by T both the central limit operator $T(\mu) = (\mu * \mu) \circ S_{\frac{1}{\sqrt{2}}}$ and the free central limit operator $T(\mu) = (\mu \boxplus \mu) \circ S_{\frac{1}{\sqrt{2}}}$. It will be clear from the context which one is meant. We denote by $\mathcal{T}$ the manifestations of T on auxiliary spaces: in Section 2, the spaces of sequences; in Section 3, the spaces of continuous functions. Precise definitions will be given at appropriate times. Also, we denote by χ the appropriate normal distributions: standard Gaussian in the classical context and the standard Wigner semicircle law [Voi85] in the free context.

2.2. **Background.** For a measure μ, its n-th moment is $m_n^\mu = \int x^n d\mu(x)$. In this section we consider only probability measures whose moments of all orders are finite. In fact, throughout most of the section we disregard the non-uniqueness

and identify the measure with its collection of moments. Thus let $\mathcal{M}$ be the space of all one-sided real-valued sequences. We will call the elements of $\mathcal{M}$ the moment sequences and denote them by $\mathbf{m} := \{m_i\}_{i=1}^{\infty}$, even though only some of them are moment sequences of measures. On $\mathcal{M}$ we set up the topology of entrywise convergence; this is the weak*-topology on the space $\mathcal{M}$ as the dual of the space of the "eventually 0" sequences, and it turns $\mathcal{M}$ into a topological vector space. Note that if a sequence of elements of $\mathcal{M}$ do in fact correspond to measures, and if its limit corresponds to a *unique* measure, then one has weak convergence of the corresponding measures [Dur91].

For every such moment sequence $\mathbf{m}$ there is also the corresponding *free cumulant* sequence $\mathbf{c} := \{c_i\}_{i=1}^{\infty}$ [Spe94, Nic95] determined by

$$(1) \qquad \text{(free)} \qquad m_k = \sum_{\substack{\pi \in \mathcal{P}_{nc}(k) \\ \pi = \{B_1, \dots, B_n\}}} \prod_{j=1}^{n} c_{|B_j|}$$

Here $\mathcal{P}_{nc}(k)$ is the set of *noncrossing* partitions of the set $\{1, \dots, k\}$ [Kre72, Spe90, Spe94, Nic95], which can be described as follows: these are partitions of the vertices of an k-gon such that the vertices in each class can be connected by lines inside the k-gon so that the lines for different classes do not cross. Also, B_i-s denote the classes of the partition π, and $|B_i|$ denotes the number of elements of B_i.

The classical cumulant sequence can also be described by a similar formula [Shi96, Nic95], namely

$$(2) \qquad \text{(classical)} \qquad m_k = \sum_{\substack{\pi \in \mathcal{P}(k) \\ \pi = \{B_1, \dots, B_n\}}} \prod_{j=1}^{n} (|B_j| - 1)! \, c_{|B_j|}$$

where $\mathcal{P}(k)$ is the collection of all partitions of $\{1, \dots, k\}$.

Let us denote the transformation from the moment sequence to the cumulant sequence determined by the formula (1) (resp., (2)) by $\mathcal{R} : \{m_i\}_{i=1}^{\infty} \to \{c_i\}_{i=1}^{\infty}$. We call $\mathcal{R}^{-1}$ the *cumulant-moment* transform. Note that $\mathcal{R}$ is given implicitly; there are also explicit formulas [Nic95]. Note also that k-th moment depends only on the cumulants of orders less than or equal to k, and vice versa. Therefore both $\mathcal{R}$ and $\mathcal{R}^{-1}$ are continuous bijections $\mathcal{M} \to \mathcal{M}$; however, we will think of the domain of $\mathcal{R}$ as moment sequences and of its range as cumulant sequences.

The point of the transformation from the moment to the cumulant sequence is that for sequences corresponding to probability measures, the appropriate action of the operator T on the cumulant side is linear. Indeed,

$$c_k^{\mu \boxplus \nu} = c_k^{\mu} + c_k^{\nu} \qquad \text{and} \qquad c_k^{\mu \circ S_r} = r^k c_k^{\mu}$$

where c^{η} are the cumulants of the measure η. That is,

$$(3) \qquad c_k^{T\mu} = 2^{1 - \frac{k}{2}} c_k^{\mu}$$

Thus define, on the space of cumulant sequences, the operator $\mathcal{T}^{\mathcal{R}}$ by $(\mathcal{T}^{\mathcal{R}}(\mathbf{c}))_k = 2^{1-k/2}c_k$, and on $\mathcal{M}$ the operator $\mathcal{T} = \mathcal{R}^{-1}{\circ}\mathcal{T}^{\mathcal{R}} \circ \mathcal{R}$. Clearly, since $\mathcal{T}^{\mathcal{R}}$ is linear, in order to linearize the operator $\mathcal{T}$, we are interested in the linearization of the cumulant-moment transform $\mathcal{R}^{-1}$.

In the sequel, by a *linearization* of a map A at a point x we mean its Gâteaux derivative: $(D_x A)(y) = \lim_{\varepsilon \to 0} \frac{A(x+\varepsilon y)-A(x)}{\varepsilon}$ when the limit exists in the appropriate topology. Note also that from (3), a fixed point of T for which the cumulant sequence is defined must have all the cumulants other than the second one equal to 0. In the classical case, this describes the Gaussian distributions; in the free case, this describes the free normal distributions, which are the dilations of the Wigner semicircle law [Spe90].

2.3. Proposition. *The linearization of the cumulant-moment transform at (the cumulant series corresponding to) the normal distribution χ (respectively, standard Gaussian in the classical case and standard Wigner semicircular distribution in the free case [VDN92]) is given by a (formal) infinite lower-triangular matrix $A = (a_{ij})_{i,j=1}^{\infty}$, where*

   a. *In the classical case, $a_{n+2k,n}$ is $(n-1)!$ times the number of partitions of $(n+2k)$ elements into classes exactly one of which contains n elements and the remaining k classes are pairs.*

   b. *In the free case, $a_{n+2k,n}$ is the number of noncrossing partitions of $(n+2k)$ elements into classes exactly one of which contains n elements and the remaining k classes are pairs.*

In both cases $a_{ij} = 0$ if $j > i$ or $(i-j)$ is odd. For explicit values, see Theorem 2.4.

Proof. The value of the n-th cumulant is a polynomial function of the first n moments only, and vice versa. Thus in the topology of entrywise convergence, the differentials of both $\mathcal{R}$ and $\mathcal{R}^{-1}$ exist.

Given two moment sequences $\mathbf{m}^o, \mathbf{m}^d$, we define the sequence $\{f(\mathbf{m}^o, \mathbf{m}^d)_n\}_{n=1}^{\infty}$ recursively by

$$(4) \qquad m_k^d = \sum_{\substack{\pi \in \mathcal{P}_{nc}(k) \\ \pi = \{B_1, \dots, B_n\}}} \sum_{i=1}^{n} \left(\prod_{j \neq i} c_{|B_j|}^o \right) f(\mathbf{m}^o, \mathbf{m}^d)_{|B_i|}$$

where $\{c_i^o\} = \mathcal{R}(\mathbf{m}^o)$ are the free cumulants. Then

$$\sum_{\substack{\pi \in \mathcal{P}_{nc}(k) \\ \pi = \{B_1, \dots, B_n\}}} \prod_{i=1}^{n} \left(c_{|B_i|}^o + \varepsilon f(\mathbf{m}^o, \mathbf{m}^d)_{|B_i|} \right) = m_k^o + \varepsilon m_k^d + o(\varepsilon)$$

Note that if $\mathbf{m}^o = \mathbf{m}^{\mu}, \mathbf{m}^d = \mathbf{m}^{\nu}$, then the last expression above is just $\mathbf{m}^{\mu+\varepsilon\nu} + o(\varepsilon)$. So the sequence $\{f(\mathbf{m}^o, \mathbf{m}^d)_n\}_{n=1}^{\infty}$ is the derivative of the moment-cumulant transform $\mathcal{R}$ at $\mathbf{m}^o$ in the direction $\mathbf{m}^d$. For $\mathbf{m}^o = \mathbf{m}^{\chi}$, the free standard normal

(semicircular) distribution, $c_i^\chi = \delta_{i2}$, and so (4) becomes

$$
(5) \qquad m_k^d = \sum_{n=1}^{k} a_{k,n} f(\mathbf{m}^d)_n
$$

where $f(\mathbf{m}^d) := f(\mathbf{m}^\chi, \mathbf{m}^d)$ and $a_{n+2k,n}$ is the number of noncrossing partitions of $(n + 2k)$ elements into classes exactly one of which has n elements and the remaining k classes are pairs. Thus $m = Af$, where A is the lower-triangular matrix $(A)_{i,j} = a_{i,j}$.

In the classical case, we start with

$$
(6) \qquad m_k^d = \sum_{\substack{\pi \in \mathcal{P}(k) \\ \pi = \{B_1, \ldots, B_n\}}} (|B_i| - 1)! \sum_{i=1}^{n} \left(\prod_{j \neq i} c_{|B_j|}^o \right) f(\mathbf{m}^o, \mathbf{m}^d)_{|B_i|}
$$

and by the same sort of reasoning see that the derivative of $\mathcal{R}$ at the standard Gaussian is given by the lower-triangular matrix A with $a_{n+2k,n} = (n-1)! \times$ (the number of partitions of $(n+2k)$ elements into classes exactly one of which contains n elements and the remaining k classes are pairs). $\qquad \square$

As stated above, the operator $\mathcal{T}^\mathcal{R}$ is linear. It is easy to see that its spectrum is discrete. Its eigenvectors are the cumulant sequences $\xi_j = \{\delta_{ij}\}_{i=1}^{\infty}$, for $j = 1, 2, \ldots$, with corresponding eigenvalues $2^{1-j/2}$. Therefore for the linearization of operator $\mathcal{T}$, the eigenvalues are the same, and the eigenvectors are the moment sequences $e_j = \{a_{i,j}\}_{i=1}^{\infty}$, where $a_{i,j}$-s are defined in the above theorem. In fact, these are true moment sequences, and so give the eigenfunctions for the central limit operator $\mathcal{T}$.

2.4. Theorem. *On the space of measures with all moments finite, the linearization of the operator T has eigenvalues $2^{1-n/2}, n = 1, 2, \ldots$. The corresponding eigenfunctions are absolutely continuous with respect to the Lebesgue measure, with densities:*

a. *In the classical case, $\frac{d^n}{dx^n} e^{-x^2/2} = e^{-x^2/2} H_n(x)$, multiples of the Hermite polynomials [Sin92].*

b. *In the free case, $\mathbf{1}_{[-2,2]}(t) \frac{1}{\sqrt{4-t^2}} T_n(t/2)$, multiples of the Chebyshev polynomials of the first kind.*

Proof. In the classical case, $a_{n+2k,n} = (n-1)! \times$ the number of partitions of $(n+2k)$ objects into one class of n elements and k classes of 2 elements. It is easy to see that $a_{n+2k,n} = \frac{(n+2k)!}{nk!2^k}$ and $a_{k,n} = 0$ for $k < n$ or $(k-n)$ odd. Therefore for fixed n the Fourier transform (defined by $\sum_{j=0}^{\infty} \frac{1}{j!} m_j (it)^j = \int e^{itx} d\mu(x)$) of the n-th eigenfunction of T is

$$
\sum_{k=0}^{\infty} a_{n+2k,n} \frac{1}{(n+2k)!} (it)^{n+2k} = \frac{1}{n} (it)^n \exp(-t^2/2)
$$

and the sum converges absolutely. Thus the eigenfunctions are the multiples of Hermite polynomials $\frac{d^n}{dx^n}e^{-x^2/2}\,dx = e^{-x^2/2}H_n(x)\,dx$ (note that these are not exactly what one usually means by the Hermite functions).

In the free case, $a_{n+2k,n} =$ the number of noncrossing partitions of $(n+2k)$ objects into one class of n elements and k classes of 2 elements. It has been calculated by Kreweras [Kre72] to be $a_{n+2k,n} = \binom{n+2k}{k}$ (one uses an inductive argument based on the following fact: a partition π with a class of n elements is noncrossing iff each of the n intervals in the complement of this class is a union of complete classes of π, and π restricted to each of these intervals is noncrossing). The Cauchy transform (defined by $\sum_{j=0}^{\infty} m_j z^{-(j+1)} = \int \frac{d\mu(x)}{z-x}$) ([Akh65, VDN92], see also the next section) of the n-th eigenfunction is $\sum_k \binom{n+2k}{k} z^{-(n+2k+1)}$. For $z \in \mathbb{C}^+$, the series converges absolutely for $|z| > 2$. Its integral is

$$F_n(z) = -\sum_k \frac{1}{n+k}\binom{n+2k-1}{k} z^{-(n+2k)} = -\sum_k \frac{1}{n+2k}\binom{n+2k}{k} z^{-(n+2k)}$$

In particular, for $n = 1$ we have $F_1(z) = -\sum_k \frac{1}{k+1}\binom{2k}{k} z^{-(2k+1)}$. Thus $F_1(z)^2 = -F_1(z)z - 1$. Therefore F_1 is related to the generating function for the Catalan numbers [Rio68], and is in fact $\frac{-z+\sqrt{z^2-4}}{2}$. Similarly

Lemma. *For $n \geq 1$ the integral of the Cauchy transform of the n-th eigenfunction is $-\frac{1}{n}\left(\frac{z-\sqrt{z^2-4}}{2}\right)^n$, i.e. $-\sum_k \frac{1}{n+2k}\binom{n+2k}{k} z^{-(n+2k)} = -\frac{1}{n}\left(\frac{z-\sqrt{z^2-4}}{2}\right)^n$*

Proof of the Lemma. The series converges absolutely for $|z| > 2$. The proof is by induction, using the identity $\left(\frac{z-\sqrt{z^2-4}}{2}\right)^n \left(\frac{z-\sqrt{z^2-4}}{2}\right)^m = \left(\frac{z-\sqrt{z^2-4}}{2}\right)^{(n+m)}$. By equating coefficients, we have to prove the combinatorial identity

$$\sum_{\substack{k+l=t \\ k,l\geq 0}} \frac{n}{n+2k}\binom{n+2k}{k}\frac{m}{m+2l}\binom{m+2l}{l} = \frac{n+m}{n+m+2t}\binom{n+m+2t}{t}$$

for $m, n = 1, 2, \ldots$ and $t = 0, 1, \ldots$. But this is a particular case of the generalized Vandermonde (also known as Rothe) identity ([Rio68, Sec. 4.5], see also [GK66]).

The above moments determine a unique distribution [Dur91], and we can see directly that the n-th eigenfunction is related to the Chebyshev polynomials of the first kind, namely it is a scalar multiple of $\mathbf{1}_{[-2,2]}(t)\frac{1}{\sqrt{4-t^2}}T_n(t/2)\,dt$, where $T_n(t) = \cos^{-1}(n\cos t)$. $\qquad\square$

2.5. *Remark.* Besides being eigenfunctions of the operator DT on a topological vector space, the above functions in fact form orthogonal bases in (smaller) Hilbert spaces. The Hermite functions $\frac{d^n}{dx^n}e^{-x^2/2} = e^{-x^2/2}H_n(x)$ for $n = 0, 1, \ldots$ form an orthogonal basis in $L^2(e^{x^2/2}\,dx)$, while the Chebyshev functions of the first kind $\mathbf{1}_{[-2,2]}(t)\frac{1}{\sqrt{4-t^2}}T_n(t/2)$ for $n = 0, 1, \ldots$ form an orthogonal basis in the space

$L^2(\mathbf{1}_{[-2,2]}(t)\sqrt{4-t^2}\,dt)$. Thus, in the classical case the operator DT is a compact self-adjoint operator on $L^2(e^{x^2/2}\,dx)$, while in the free case DT is a compact self-adjoint operator on $L^2(\mathbf{1}_{[-2,2]}(t)\sqrt{4-t^2}\,dt)$.

We have the following interpretation of the CLT.

2.6. Corollary. *We say that a measure μ is in $L^2(\varphi)$ if $\mu \ll \varphi\,dx$ and $\frac{d\mu}{dx} \in L^2(\varphi)$. Note that for all the measures in $L^2(e^{x^2/2}\,dx)$ or $L^2(\mathbf{1}_{[-2,2]}(t)\sqrt{4-t^2}\,dt)$, the moments of all orders are finite.*

  a. *On the space of probability distributions in $L^2(e^{x^2/2}\,dx)$, with mean 0 and variance 1, the normal distribution χ as a fixed point of the central limit operator T is strictly spectrally stable, that is, the differential of the operator at this point has the spectrum inside the unit disc.*
  b. *On the space of probability distributions in $L^2(\mathbf{1}_{[-2,2]}(t)\sqrt{4-t^2}\,dt)$, with mean 0 and variance 1, the free normal (semicircular) distribution χ as a fixed point of the free central limit operator T is strictly spectrally stable.*

Proof. Classical case: [Sin92] The Hermite functions $\frac{d^n}{dx^n}e^{-x^2/2} = e^{-x^2/2}H_n(x)$ for $n = 0, 1, \ldots$ form an orthogonal basis in $L^2(e^{x^2/2}\,dx)$. The conditions on the moments of the distribution in the hypothesis mean that in the linear approximation we consider only the perturbations f with 0th, 1st, 2nd moments equal to 0. This means precisely that f is orthogonal to $e^{-x^2/2}$ and (the densities of) the first two eigenfunctions of T. Since the eigenvalues of T are $2^{1-n/2}$, all eigenvalues for $n > 2$ are less than 1.

Free case: the Chebyshev functions $\mathbf{1}_{[-2,2]}(t)\frac{1}{\sqrt{4-t^2}}T_n(t/2)$ for $n = 0, 1, \ldots$ form an orthogonal basis in the space $L^2(\mathbf{1}_{[-2,2]}(t)\sqrt{4-t^2}\,dt)$. Again the eigenvalues are less than 1 for all but the first 3 of these. But here, the hypothesis correspond to the orthogonality to the first 3 Chebyshev functions only if all the distributions considered are *supported in the same interval* $[-2,2]$. Note, however, that due to the results in [BV95], this restriction is weaker than it appears. $\qquad\square$

Thus we would expect that on this subspace the fixed point is attracting, just as the CLT states.

2.7. *Remark.* The moments $a_{n+2k,k}$ can be calculated in a way similar to the above for the setting of the R_q transforms [Nic95], related to q-independence. However, there does not seem to be in that case a nice recurrence formula, and so the corresponding eigenfunctions are not calculated directly. Moreover, note that the orthogonality conditions in the above Corollary are different in the classical and the free cases: in the free case the inner product is given by the free normal (semicircular) distribution, while in the classical case it is the inverse of the normal distribution. Also, we are not aware of any standard interpolation between the Hermite polynomials and the Chebyshev polynomials of the first kind. Thus one would not necessarily expect to have a similar construction for the interpolations between free and classical cases, e.g. related to q-independence.

3. THE ANALYTIC APPROACH

In this section we consider the problem of linearizing the operator T by analytic means. First we briefly go over the classical situation.

3.1. **Classical Picture.** Let $\alpha \in (0,2]$, $\beta = 2^{-1/\alpha}$, $T_\alpha \mu = (\mu * \mu) \circ S_\beta$. For $\varphi_\alpha = \alpha$-strictly stable distribution [Shi96, Dur91] (the skewness coefficient does not appear explicitly in the sequel and so is not included in the notation), $T_\alpha(\varphi_\alpha) = \varphi_\alpha$. Then the differential of T_α at φ_α is

$$DT_\alpha \nu = 2(\nu * \varphi_\alpha) \circ S_\beta$$

Taking the Fourier transforms,

(7) $$\widehat{DT_\alpha \nu}(t) = 2\hat{\nu}(\beta t)\hat{\varphi}_\alpha(\beta t)$$

Also by stability $\hat{\varphi}_\alpha^2(\beta t) = \hat{\varphi}_\alpha(t)$. Therefore for $\hat{\nu}(t) = h(t)\hat{\varphi}_\alpha(t)$, the right-hand-side expression in (7) is $2h(\beta t)\hat{\varphi}_\alpha(t)$. For h, on the space of continuous functions the eigenfunctions are $h(t) = t^a$, $a \in \mathbb{C}$, $\text{Re}\, a > 0$ or $a = 0$, with eigenvalues $2 \cdot \beta^a = 2 \cdot 2^{-a/\alpha}$, corresponding to $\hat{\nu}_a(t) = t^a \hat{\varphi}_\alpha(t)$. Here we use the principal branch of the logarithm. Now let $\beta = 1/\sqrt{2}$, i.e. $\varphi_\alpha = \varphi_2 = \chi$. In this case, among all the eigenfunctions we can distinguish the integer values of a as follows: among the functions t^a, the smooth ones are precisely those for $a \in \mathbb{N}$. Thus among all ν_a, the ones whose densities decay faster than any polynomial are just the ν_n-s, $n \in \mathbb{N}$ [Shi96, Dur91]. These measures are manifestly in $L^2(e^{x^2/2}\, dx)$. They are $\nu_n = \frac{d^n}{dx^n} e^{-x^2/2}\, dx$, with eigenvalues $2^{1-n/2}$, and we obtain the result of the previous section.

3.2. **Free Picture.** In the free probability picture, the main device is the R-transform, introduced by Voiculescu [Voi86, BV93, VDN92]. Given a measure μ, for $z \in \mathbb{C} \backslash \text{supp}(\mu)$ one defines the Cauchy transform (sometimes called Stieltjes or Borel transform) of μ by $G_\mu(z) = \int \frac{d\mu(t)}{z-t}$. For positive μ the Cauchy transform is an analytic map $\mathbb{C}^+ \leftrightarrow \mathbb{C}^-$; it has the property $\bar{G}_\mu(z) = G_\mu(\bar{z})$. The measure can be reconstructed from its Cauchy transform by taking the boundary values $-\frac{1}{\pi} \text{Im}\, G_\mu(x + 0i)$ [Akh65, Ch.3, Addenda and Problems], [Hör90, 3.1].

On a nontangential neighborhood of 0 (Stolz angle) in $\mathbb{C}^-$, we can define $K_\mu(w) = G_\mu^{-1}(w)$, and the R-transform $R_\mu(w) = K_\mu(w) - \frac{1}{w}$. R_μ is an analytic map $\mathbb{C}^- \to \mathbb{C}^-$ on a nontangential neighborhood of 0. The main property of the R-transform is that it also linearizes the additive free convolution: $R_{\mu \boxplus \nu}(w) = R_\mu(w) + R_\nu(w)$. In fact, if all the moments of a measure μ are finite, then $R_\mu(w) = \sum_{i=1}^\infty c_i^\mu w^{i-1}$, where c_i are the free cumulants. Also $R_{\mu \circ S_r}(w) = r R_\mu(rw)$. Thus the action of the operator T_α on the R-transform side is just $R_{T_\alpha \mu}(w) = 2\beta R_\mu(\beta w)$, and in particular it is linear. Here we again define the operator

$$T_\alpha(\mu) = (\mu \boxplus \mu) \circ S_\beta$$

for $\alpha \in (0, 2]$, and $\varphi_\alpha =$ free α-strictly stable distribution [BV93, Pat95, BPB96]. By the above observations, the linearization of T_α is again given by the linearization of the R-transform.

The quasilinear differential equation governing the behavior of free convolution semigroups has first appeared in [Voi86, Theorem 4.3]. Here we need a variant of that theorem. The proof is quite similar to [Voi86].

3.3. Theorem. *Let ν be a freely infinitely divisible probability measure [Voi86, VDN92]. Let ψ be a function analytic in $\mathbb{C}^+$. For any point $z \in \mathbb{C}^+$, for small enough t the function $G(z, t)$, which is the functional inverse of $K_\nu(z) + t\psi(z)$, is defined at z. Consequently for all $z \in \mathbb{C}^+$,*

$$(8) \qquad G'_\nu(z)\psi(G_\nu(z)) + \frac{\partial G}{\partial t}(z, 0) = 0$$

Proof. By [BV93, Proposition 5.12] for ν freely infinitely divisible, G_ν maps $\mathbb{C}^+$ conformally onto a domain. Let Ω be a bounded domain in $\mathbb{C}^+$ whose closure is contained in $\mathbb{C}^+$. Choose t so that for $z \in \Omega$, $t\,|\psi'(G_\nu(z))| < 1$. Then the function $z + t\psi(G_\nu(z))$ is univalent on Ω and invertible on its image. Denote this image by Ω' and this inverse function by f_t. Since, for $z \in \Omega$,

$$(K_\nu + t\psi)(G_\nu(z)) = z + t\psi(G_\nu(z))$$

we also have

$$(K_\nu + t\psi)(G_\nu(f_t(z)) = z$$

for $z \in \Omega'$. Consequently we can define $G(z, t) = G_\nu(f_t(z))$ for $z \in \Omega'$.

As one possible construction, define

$$\Omega_n = \left\{ z \mid -n \leq \operatorname{Re} z \leq n, \frac{1}{n} \leq \operatorname{Im} z \leq n \right\}$$

Let t_n satisfy the condition above and also $t_n\,|\psi(G_\nu(z))| < 1/n$ for $z \in \Omega_n$. Then $G(z, t)$ is defined on

$$\Omega'_n = \left\{ z \mid -n + \frac{1}{n} \leq \operatorname{Re} z \leq n - \frac{1}{n}, \frac{2}{n} \leq \operatorname{Im} z \leq n - \frac{1}{n} \right\}$$

These domains exhaust $\mathbb{C}^+$. On their common domains, since their inverse is analytic in t, $G(z, t)$ is differentiable in t; the derivative exists as a limit in the topology of uniform convergence on compact sets in the upper half plane.

By definition $G(K_\nu + t\psi(z), t) = z$ for $z \in G_\nu(\Omega)$. Differentiating with respect to t at $t = 0$, we get

$$\frac{\partial G}{\partial z}(K_\nu(z), 0)\psi(z) + \frac{\partial G}{\partial t}(K_\nu(z), 0) = 0$$

for $z \in G_\nu(\mathbb{C}^+)$. Substituting $G_\nu(z)$ for z, we get the required equation, for $z \in \mathbb{C}^+$. $\qquad \square$

3.4. *Remark.* In the above theorem no consideration is given to the positivity or even existence of the distribution corresponding to the Cauchy transform $G(z, t)$. While we do not know of a satisfactory description of these, there are various conditions.

3.4.1. *Necessary Conditions.* For $K_\nu + \psi$ to correspond to a positive measure, it is necessary that (1) the nontangential limit of $z\psi(z)$ as $z \to 0$ be 0, and (2) there exist a Stolz angle at 0, $\Gamma \subset \mathbb{C}^+$ s.t. $(K_\nu + \psi)(\Gamma) \subset \mathbb{C}^-$ [BV93].

3.4.2. *Sufficient Conditions.*
  a. The following are the known cases where $K_\nu + \psi$ corresponds to a positive measure: (1) ψ is an R-transform of a positive (hence necessarily freely infinitely divisible) measure [Voi86]. (2) ν is the free normal (semicircular) distribution; ψ is analytic in a neighborhood of the unit disc, sufficiently small, and $\psi(\bar{z}) = \bar{\psi}(z)$ [BV95].
  b. For a measure μ and a distribution ν, the R-transform of $((1 - \varepsilon)\mu + \varepsilon\nu)$ is

$$R_{(1-\varepsilon)\mu+\varepsilon\nu} = R_\mu(w) - \varepsilon K'_\mu(w) \cdot (G_\nu - G_\mu)(K_\mu(w)) + o(\varepsilon)$$

Denote $\psi = K'_\mu(w) \cdot (G_\nu - G_\mu)(K_\mu(w))$. Then if ν is positive, the deformation in the direction of ψ is tangent to a curve of positive measures. Note also that we have the inverse formula,

$$G_\nu(w) = G_\mu(w) + \psi(G_\mu(w)) \cdot G'_\mu(w)$$

3.5. **Discussion.** The theorem can be interpreted as follows. Let G be a univalent function on $\mathbb{C}^+$. Define $K = G^{-1}$, $\mathcal{R}(G)(z) = K(z) - 1/z$, and let ψ be analytic on $G(\mathbb{C}^+)$. Then the derivative of the map $\mathcal{R}^{-1}$ (which is nothing other than the differential of the operation of functional inversion) at a point $\mathcal{R}(G)$ in the direction ψ is $-G'\psi(G)$. This linear map is invertible; the inverse linear map (which is the differential of $\mathcal{R}$) is $\psi \mapsto -K'\psi(K)$ (where ψ is now analytic in $\mathbb{C}^+$). Finally, let $\mathcal{T}^{\mathcal{R}}_\alpha(\psi)(z) = 2\beta\psi(\beta z)$ and denote $G_\alpha := G_{\varphi_\alpha}$. Then the derivative of $\mathcal{T}_\alpha = \mathcal{R}^{-1} \circ \mathcal{T}^{\mathcal{R}}_\alpha \circ \mathcal{R}$ at G_α is

$$D_{G_\alpha}\mathcal{T}_\alpha(\psi)(z) = 2G'_\alpha(z)\beta K'_\alpha(\beta G_\alpha(z))\psi(K_\alpha(\beta G_\alpha(z)))$$
$$= 2\psi(\omega_\alpha(z))\omega'_\alpha(z)$$

Here

$$\omega_\alpha(z) = K_{\varphi_\alpha}(\beta G_{\varphi_\alpha}(z)) = \frac{1}{\beta}\omega_{\varphi_\alpha \circ S_\beta, \varphi_\alpha}$$
$$= \frac{1}{\beta}\omega_{\varphi_\alpha \circ S_\beta, \varphi_\alpha \circ S_\beta \boxplus \varphi_\alpha \circ S_\beta} = \frac{1}{\beta}K_{\varphi_\alpha \circ S_\beta} \circ G_{\varphi_\alpha \circ S_\beta \boxplus \varphi_\alpha \circ S_\beta}$$

is a particular instance of the transition probability function of [Voi93, Bia98].

The eigenfunctions of $\mathcal{T}^{\mathcal{R}}_\alpha$ on the space of all analytic functions in the upper half plane are of the form $te^{i\phi}z^a$, with eigenvalues $2^{1-(a+1)/\alpha}$. Restricting to various spaces selects particular values of ϕ, a. Thus the eigenfunctions for the differential of $\mathcal{T}_\alpha$ (resp., T_α) are the (boundary values of) the functions $e^{i\phi}G'_\alpha G^a_\alpha$.

3.6. Example. For the free normal (semicircular) case $\alpha = 2$, $\nu = \chi$ the Cauchy transforms of the eigenfunctions of the operator DT (which are the eigenfunctions of the operator $D\mathcal{T}$) are given by

$$G'_\chi G^a_\chi = e^{i\phi} \frac{1}{\sqrt{z^2 - 4}} \left(\frac{z - \sqrt{z^2 - 4}}{2} \right)^{x + yi}$$

By taking the boundary values $-\frac{1}{\pi} \operatorname{Im} G(t + 0i)$ (see [Akh65, Hör90]), the eigenfunctions themselves are

$$\frac{1}{\sqrt{t^2 - 4}} \left| \frac{t - \sqrt{t^2 - 4}}{2} \right|^x e^{-y\pi} \sin\left(y \log \left| \frac{t - \sqrt{t^2 - 4}}{2} \right| + \phi + x\pi \right) \mathbf{1}_{(-\infty, -2]}(t)\, dt$$

$$+ \frac{1}{\sqrt{t^2 - 4}} \left| \frac{t - \sqrt{t^2 - 4}}{2} \right|^x e^{-y\pi} \sin\left(y \log \left| \frac{t - \sqrt{t^2 - 4}}{2} \right| + \phi \right) \mathbf{1}_{[2, \infty)}(t)\, dt$$

$$+ \frac{1}{\sqrt{4 - t^2}} \exp\left(y \cos^{-1}(t/2) \right) \cos\left(x \cos^{-1}(t/2) - \phi \right) \mathbf{1}_{[-2, 2]}(t)\, dt$$

It is easy to see that the Criterion 3.4.1 requires that (for some ϕ) we have $x = \operatorname{Re} a \geq 1$ or ($x = \operatorname{Re} a \geq -1$, $y = \operatorname{Im} a = 0$); the corresponding point spectrum is the union of the unit disc and the interval $[1, 2]$. On the other hand, for $-1 < x < 1$, $y = 0$ the functions are the R-transforms of freely stable distributions, while for $x \in \mathbb{N}$, $y = 0$ the functions are entire. Thus by Criterion 3.4.2a the corresponding eigenfunctions are in the tangent space to the space of positive measures.

All the moments of a measure are finite iff its Cauchy transform is analytic at infinity. If a function G is defined by the above expression on $\mathbb{C}^+$ and satisfies $G(\bar{z}) = \overline{G(z)}$, it is analytic at infinity iff $a \in \mathbb{N}$ and $\phi = 0$. Notice that among the above measures, these are precisely the compactly supported ones. Explicitly their Cauchy transforms are $-\frac{1}{\sqrt{z^2 - 4}} \left(\frac{z - \sqrt{z^2 - 4}}{2} \right)^n$. The eigenfunctions themselves are

$$\frac{1}{\sqrt{4 - x^2}} \cos(n \cos^{-1}(x/2)) \mathbf{1}_{[-2, 2]}(x)\, dx$$

That is, we recover the Chebyshev functions of the first kind.

3.7. *Remark.* Here we can see another difference from the classical case. As noted above, by a result of Bercovici and Voiculescu [BV95] the deformations in the directions z^n, $n \in \mathbb{N}$ actually produce positive measures (for small enough time). This is in contrast with a classical theorem of Marcinkiewicz, which states that for P a polynomial, e^P is never a characteristic function (i.e. a Fourier transform of a positive measure) if the degree of P is greater than 2 (see e.g. [Ram67, Thm 3.13]).

3.8. Example. For the 1-stable symmetric distribution, which is the Cauchy distribution φ_1, the eigenfunction Cauchy transforms are $\frac{1}{(z - i)^a}$, and the eigenfunctions are $D^a \frac{1}{x^2 + 1}\, dx$. Notice that these are exactly the same as in the classical

case. Therefore not only are the free 1-stable distributions the same as classical ones [BV93], but their small neighborhoods look the same as well.

3.9. *Remark.* For a general freely stable distribution, using for example the formula $D_G \mathcal{R}^{-1}(\psi) = -\frac{\psi}{K'}(G)$, one can obtain parametric expression à la Biane [BPB96, Appendix] for the densities of the corresponding boundary values. In particular, one has such expressions for the densities of the eigenfunctions of the free stable central limit operators. It is not clear whether they are of use.

3.10. **Composition operators.** Finally, we have a brief discussion of the connections with the theory of composition operators (see e.g. [Val31, CM95]). The action of the operator $\mathcal{T}_\alpha$ on the primitive (in $\mathbb{C}^+$) of a Cauchy transform is, up to a multiplicative constant 2 and up to an additive constant, just the composition with the function ω_α.

One of the main theorems about composition operators is that such an operator is necessarily conjugate to an operator of composition with a linear function [CM95, Thm. 2.53]. In our case, in the terminology of [CM95, Section 2.4] the operator $\mathcal{T}_\alpha$ has a natural halfplane-dilation model provided by conjugating with the linearization of the R-transform, i.e. $G_\alpha \circ \omega_\alpha = \beta G_\alpha$. A *fundamental set* [CM95, Defn. 2.54] for ψ is an open, connected, simply connected domain Δ such that $\psi(\Delta) \subset \Delta$ and the iterates of any compact set end up in it after a finite number of steps. It is not hard to see (e.g. [BPB96, Appendix]) that $\mathbb{C}^+$ serves as a fundamental set for both G_α and ω_α while $G(\mathbb{C}^+)$ is a fundamental set for K_α.

A number of results on the spectra of composition operators on various classical spaces are known. In particular, on a Hardy space H^∞ the spectral radius of a composition operator is equal to 1 [CM95, 3.1]. In our case the composition operator is defined on tangent spaces to a certain cone in H^∞.

References

[Akh65] N.I. Akhiezer, *The Classical Moment Problem and some Related Questions in Analysis*, Hafner Publishing Co., New York, 1965.

[Bar86] A.R. Barron, Entropy and the central limit theorem, Ann. Probab. **14** (1986), no. 1, 336–342.

[BPB96] H. Bercovici and V. Pata, Stable laws and domains of attraction in free probability theory, with an Appendix: P. Biane, The density of free stable distributions, preprint (1996).

[BV93] H. Bercovici and D.-V. Voiculescu, Free convolution of measures with unbounded support, Indiana Univ. Math. J. **42** (1993), no. 3, 733–773.

[BV95] _____, Superconvergence to the central limit and failure of the Cramér theorem for free random variables, Probab. Theory Related Fields **103** (1995), no. 2, 215–222.

[Bia98] P. Biane, Processes with free increments, Math. Z. **227** (1998), no. 1, 143–174.

[CM95] C.C. Cowen and B.D. MacCluer, *Composition Operators on Spaces of Analytic Functions*, CRC Press, Boca Raton, FL, 1995.

[Dur91] R. Durrett, *Probability*, Wadsworth & Brooks/Cole, Pacific Grove, CA, 1991.

[GK66] H.W. Gould and J. Kaucký, Evaluation of a class of binomial coefficient summations, J. Combinatorial Theory **1** (1966), 233–247.

[Gol76] J.A. Goldstein, Semigroup-theoretic proofs of the central limit theorem and other theorems of analysis, Semigroup Forum **12** (1976), no. 3, 189–206.

[Hör90] L. Hörmander, *The Analysis of Linear Partial Differential Operators. I*, 2nd ed., Grundlehren der Mathematischen Wissenschaften, vol. 256, Springer-Verlag, Berlin, 1990, Distribution theory and Fourier analysis.

[HW84] G.G. Hamedani and G.G. Walter, A fixed point theorem and its application to the central limit theorem, Arch. Math. (Basel) **43** (1984), no. 3, 258–264.

[Kre72] G. Kreweras, Sur les partitions non croisées d'un cycle, Discrete Math. **1** (1972), no. 4, 333–350.

[Maa92] H. Maassen, Addition of freely independent random variables, J. Funct. Anal. **106** (1992), no. 2, 409–438.

[Nic95] A. Nica, A one-parameter family of transforms, linearizing convolution laws for probability distributions, Comm. Math. Phys. **168** (1995), no. 1, 187–207.

[Pat95] V. Pata, Lévy type characterization of stable laws for free random variables, Trans. Amer. Math. Soc. **347** (1995), no. 7, 2457–2472.

[Ram67] B. Ramachandran, *Advanced Theory of Characteristic Functions*, Statistical Publishing Society, Calcutta, 1967.

[Rio68] J. Riordan, *Combinatorial Identities*, John Wiley & Sons Inc., New York, 1968.

[Shi96] A.N. Shiryaev, *Probability*, 2nd ed., Graduate Texts in Mathematics, vol. 95, Springer-Verlag, New York, 1996.

[Sin76] Ya.G. Sinaĭ, Self-similar probability distributions, Teor. Verojatnost. i Primenen. **21** (1976), no. 1, 63–80.

[Sin92] _____, *Probability Theory*, Springer Textbook, Springer-Verlag, Berlin, 1992.

[Spe90] R. Speicher, A new example of "independence" and "white noise", Probab. Theory Related Fields **84** (1990), no. 2, 141–159.

[Spe94] _____, Multiplicative functions on the lattice of noncrossing partitions and free convolution, Math. Ann. **298** (1994), no. 4, 611–628.

[Swa91] S. Swaminathan, Fixed-point-theoretic proofs of the central limit theorem, in *Fixed Point Theory and Applications* (Marseille, 1989), Pitman Res. Notes Math. Ser., vol. 252, Longman Sci. Tech., Harlow, 1991, pp. 391–396.

[Tro59] H.F. Trotter, An elementary proof of the central limit theorem, Arch. Math. **10** (1959), 226–234.

[Val31] G. Valiron, Sur l'iteration des fonctions holomorphes dans un demi-plan, Bull. Sci. Math. (2) **55** (1931), 105–128.

[Voi85] D.-V. Voiculescu, Symmetries of some reduced free product C^*-algebras, in *Operator Algebras and their Connections with Topology and Ergodic Theory* (Buşteni, 1983), Lecture Notes in Math., vol. 1132, Springer, Berlin, 1985, pp. 556–588.

[Voi86] _____, Addition of certain noncommuting random variables, J. Funct. Anal. **66** (1986), no. 3, 323–346.

[Voi93] _____, The analogues of entropy and of Fisher's information measure in free probability theory. I, Comm. Math. Phys. **155** (1993), no. 1, 71–92.

[VDN92] D.-V. Voiculescu, K.J. Dykema, and A. Nica, *Free random variables*, CRM Monograph Series, vol. 1, American Mathematical Society, Providence, RI, 1992.

DEPARTMENT OF MATHEMATICS, UNIVERSITY AT CALIFORNIA, BERKELEY, CA 94720, USA

E-mail address: mashel@math.berkeley.edu